

APLIKASI METODE *SUPPORT VECTOR MACHINE* (SVM)

PADA DATA BIOINFORMATIKA

(Studi Kasus: Klasifikasi *Genes Expression*

Data Maize Nested Association Mapping)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana

Jurusan Statistika



Ummi Maftuhatul Mu'inah

14611202

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS ISLAM INDONESIA

YOGYAKARTA

2018

HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR

Judul : Aplikasi Metode *Support Vector Machine* (SVM) pada
Data Bioinformatika
(Studi Kasus : Klasifikasi *Genes Expression* Data *Maize*
NAM)

Nama Mahasiswa : Ummi Maftuhatul Mu'inah

Nomor Mahasiswa : 14611202

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Yogyakarta, 1 Agustus 2018

Menyetujui,

Dosen Pembimbing I

Dr. techn. Rohmatul Fajriyah, S.Si., M.Si

Menyetujui,

Dosen Pembimbing II

Husna Nugrahapraja, Ph.D

HALAMAN PENGESAHAN

TUGAS AKHIR

APLIKASI METODE *SUPPORT VECTOR MACHINE* (SVM)
PADA DATA BIOINFORMATIKA

Nama Mahasiswa : Ummi Maftuhatul Mu'inah

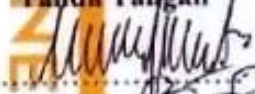
Nomor Mahasiswa : 14 611 202

TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL 1 AGUSTUS 2018

Nama Penguji

1. Muryanto, S.P., M.Sc
2. Arum Handini Primandari, S.Pd.Si., M.Sc
3. Husna Nugrahaputra, Ph.D
4. Dr. techn. Rohmatul Fajriyah, S.Si., M.Si

Tanda Tangan


.....

.....

.....

.....

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Prof. Muryanto, S.Pd., M.Si., Ph.D)

KATA PENGANTAR



Assalamu'alaikum Wr. Wb

Puji syukur penulis panjatkan kehadiran Allah SWT. yang senantiasa melimpahkan rahmat serta hidayah-Nya berupa keimanan, kesehatan dan kelancaran sehingga penulis dapat menyelesaikan tugas akhir yang berjudul **“Aplikasi Metode *Support Vector Machine* (SVM) pada Data Bioinformatika (Studi Kasus: Klasifikasi *Genes Expression Data Maize NAM*)”**. Shalawat serta salam selalu tercurah kepada Nabi Muhammad SAW. beserta keluarga, sahabat dan pengikut beliau hingga akhir zaman.

Tugas Akhir ini disusun untuk memenuhi syarat untuk menyelesaikan Program Sarjana Strata Satu (S1) Jurusan Statistika pada FMIPA Universitas Islam Indonesia. Penulis menyadari tanpa adanya bimbingan dari berbagai pihak, penyusunan tugas akhir ini tidak dapat terselesaikan dengan baik. Oleh karena itu penulis juga menyampaikan terima kasih kepada :

1. Orang tua dan keluarga yang selalu memberikan do'a dan dukungan positif untuk penulis.
2. Prof. Riyanto, S.Pd., M.Si., Ph.D, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia berserta jajarannya.
3. Dr. RB. Fajriya Hakim, M.Si., selaku Ketua Jurusan Statistika, Universitas Islam Indonesia beserta jajarannya.
6. Ibu Dr. Techn. Rohmatul Fajriyah, S.Si., M.Si., selaku Dosen Pembimbing I atas waktu, tenaga, motivasi, ilmu, nasehat, serta bimbingannya sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik.
7. Bapak Husna Nugrahapraja, Ph.D., selaku Dosen Pembimbing II atas ilmu, tenaga, waktu, nasehat, serta bimbingannya sehingga tugas akhir ini dapat selesai dengan baik.
8. Seluruh staf pengajar Jurusan Statistika Universitas Islam Indonesia terimakasih atas ilmu yang diberikan, semoga senantiasa diberikan kebaikan oleh Allah SWT.

9. Teman-teman SBRC : Ika, Lina, Hawila, Leni, Himelda, Nanda, Maidah, Husain, Ari, Shodiq, Gilang, Aat, dan Mas Rakhmat yang selalu memberikan bantuan, dukungan, do'a dan saran-saran selama pelaksanaan dan penyusunan tugas akhir.
10. Teman-teman seperjuangan Statistika UII Angkatan 2014, sukses dan berjuanglah menuju hari esok yang lebih baik.
11. Semua pihak yang terlibat dalam penyusunan tugas akhir ini yang tidak bisa disebutkan satu per satu yang telah memberikan dukungan dan motivasi selama menyelesaikan tugas akhir ini.

Semoga segala bantuan dan dukungan yang telah diberikan mendapat balasan yang lebih banyak dari Allah SWT. Akhir kata penulis harapkan adanya kritik dan saran yang membangun demi kesempurnaan penyusunan dan penulisan tugas akhir ini. Semoga tugas akhir ini dapat berguna dan bermanfaat bagi penulis pada khususnya dan bagi pembaca pada umumnya.

Wassalamu'alaikum Wr. Wb.

Yogyakarta, Juli 2018

Ummi Maftuhatul Mu'inah

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN PEMBIMBING	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR.....	v
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR LAMPIRAN	xi
PERNYATAAN.....	xii
INTISARI	xiii
ABSTRACT	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
BAB II KAJIAN PUSTAKA	6
2.1 Penelitian dengan <i>Support Vector Machine</i>	6
2.2 Penelitian Terkait Jagung	7
2.3 Penelitian Terkait SVM di Bioinformatika	8
BAB III LANDASAN TEORI.....	11
3.1 Bioinformatika	11
3.2 <i>Next Generation Sequencing</i> (NGS).....	12
3.3 DNA dan <i>gene Expression</i>	13
3.4 <i>Filtering</i>	14
3.5 Klasifikasi	15
3.6 <i>Support Vector Machine</i> (SVM).....	16
3.7 <i>Soft Margin</i>	18
3.8 <i>SVM Multiclass</i>	18
3.9 <i>Confusion Matrix</i>	21
3.10 <i>Receiver Operating Characteristic</i> (ROC) <i>Curve</i>	23
3.11 Jagung (<i>Zea Mays</i>).....	24
BAB 4 METODOLOGI PENELITIAN.....	26
4.1 Jenis dan Sumber Data.....	26
4.2 Tempat dan Waktu Penelitian.....	26
4.3 Variabel Penelitian	26

4.4 Metode Analisis Data	26
4.5 Tahapan Penelitian.....	27
BAB 5 HASIL DAN PEMBAHASAN	29
5.1 Deskripsi Data	29
5.2 Analisis Data Bioinformatika	30
5.2.1 <i>Filtering</i>	30
5.2.2 Klasifikasi SVM	31
5.2.3 <i>Confusion Matrix</i>	32
5.3 Model Klasifikasi SVM.....	35
BAB 6 PENUTUP.....	37
6.1 Kesimpulan	37
6.2 Saran	38
DAFTAR PUSTAKA	39
LAMPIRAN.....	42

DAFTAR GAMBAR

Gambar 3.1 Mencari <i>Hyperplane</i> Terbaik	16
Gambar 3.2 <i>Soft Margin</i>	18
Gambar 3.3 Contoh klasifikasi metode <i>One-Against-All</i> untuk 4 Kelas	20
Gambar 3.4 Contoh klasifikasi metode <i>One-Against-One</i> untuk 4 Kelas	21
Gambar 4.1 <i>Flowchart</i> tahapan penelitian	27
Gambar 4.2 <i>Flowchart</i> tahapan analisis data	28
Gambar 5.1 <i>Stripchat</i> Ekspresi Gen Zm00001d027230	30

DAFTAR TABEL

Tabel 3.1 Contoh SVM <i>one-against-all</i> untuk 4 kelas	20
Tabel 3.2 Contoh SVM <i>one-against-one</i> untuk 4 kelas	21
Tabel 3.3 <i>Confusion matrix</i> dimensi 2x2	22
Tabel 3.4 <i>Confussion matrix multiclass</i>	23
Tabel 3.5 <i>Classifying the accuracy of diagnostic test</i>	24
Tabel 5.1 Jumlah organ data <i>Maize NAM</i>	29
Tabel 5.2 <i>Confusion matrix</i> data latih	32
Tabel 5.3 <i>Confusion matrix</i> data uji	33
Tabel 5.4 Bobot <i>gene names</i>	35

DAFTAR LAMPIRAN

Lampiran 1. <i>Syntax Input Data</i>	42
Lampiran 2. <i>Syntax Memanggil Data yang digunakan</i>	42
Lampiran 3. <i>Syntax Memanggil Kolom Gene Names, Chromosome, dan Start</i> ..	42
Lampiran 4. <i>Syntax Mendefinisikan Kelas Klasifikasi</i>	42
Lampiran 5. <i>Syntax Filtering</i>	43
Lampiran 6. <i>Syntax Analisis SVM</i>	44
Lampiran 7. <i>Syntax Menyusun Data Set</i>	44
Lampiran 8. <i>Syntax Tuning</i>	44
Lampiran 9. <i>Syntax Confusion Matrix</i>	45
Lampiran 10. <i>Syntax Kurva ROC dan AUC</i>	45
Lampiran 11. <i>Syntax Menghitung Bobot</i>	46
Lampiran 12. <i>Syntax Membuat GGPlot</i>	46
Lampiran 13. <i>Syntax Membuat Stripchart</i>	46
Lampiran 14. <i>Syntax Session Info</i>	47

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, Juli 2018



Ummi Maftuhatul Mu'inah

APLIKASI METODE SUPPORT VECTOR MACHINE (SVM)

PADA DATA BIOINFORMATIKA

(Studi Kasus : Klasifikasi *Genes Expression Data Maize NAM*)

Ummi Maftuhatul Mu'inah

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Islam Indonesia

INTISARI

Bioinformatika merupakan perpaduan informatika dan ilmu biologi yang mempelajari penerapan teknik komputasi untuk mengelola dan menganalisis informasi-informasi data biologi. Bioinformatika memicu peningkatan pemahaman manusia dalam bidang genomik yang terkait dengan penggunaan sekuens Deoxyribonucleic Acid (DNA). Database genom (ekspresi dan sekuens gen) berkembang sangat pesat dan berskala besar. Pada penelitian ini, data genom yang digunakan adalah nilai ekspresi gen pada dataset Maize NAM. Tujuan dari penelitian ini untuk mengklasifikasikan informasi nilai ekspresi gen berdimensi tinggi berdasarkan pada karakteristik organ apex, ear, root, shoot, dan tassel. Salah satu metode statistik yang dapat digunakan untuk klasifikasi adalah metode Support Vector Machine (SVM). Peneliti melakukan proses filtering dengan menghapus nilai nol dan uji ANOVA untuk mengurangi kompleksitas data. Berdasarkan penelitian diperoleh nilai akurasi sebesar 95% dan nilai AUC sebesar 0,989. Hal ini menunjukkan bahwa model SVM yang terbentuk dapat mengklasifikasikan nilai ekspresi gen galur jagung berdasarkan karakteristik organ dengan sangat baik.

Kata Kunci : Bioinformatika, Klasifikasi, Jagung, Support Vector Machine, AUC.

APPLICATION OF SUPPORT VECTOR MACHINE (SVM) IN BIOINFORMATICS

(Case Study: Classification of Genes Expression on Maize NAM)

Ummi Maftuhatul Mu'inah

Department of Statistics, Faculty Mathematics and Natural Sciences

Islamic University of Indonesia

ABSTRACT

Bioinformatics is a combination of informatics and biology that studies the application of computational techniques to manage and analyze biological data information. Bioinformatics makes increased human understanding in the field of genomics related to the use of Deoxyribonucleic Acid (DNA) sequences. Genome database (expression and gene sequences) is growing very rapidly and on a large-scale. The genom data of this research was Maize NAM expression datasets. The purpose of this research was to classify the high-dimensional information on genes expression based on organs characteristics: apex, ear, root, shoot, and tassel. One of the statistical methods that can be used for classification is method of Support Vector Machine (SVM). The researcher conducted a filtering process by removing null value and ANOVA test to reduce data complexity. As the results, the accuracy value was 95% and AUC value was 0.989. It shows that SVM model can classify gene expression based on characteristics of the organs very well.

Keywords: *Bioinformatics, Classification, Maize, Support Vector Machine, AUC.*